Gauging Systematic Errors in NMME models

Suranjana Saha and Huug van den Dool
NCEP/NWS/NOAA

Acknowledgement: Qin Zhang, CPC

Predictability in NWP models

- Even when the skill of models cannot be verified, say for a weather forecast at day 18, or a seasonal forecast 13 years out, they can still be validated in terms of systematic error (SE).

- How well do state of the art models reproduce the observed climate ? Climate is the zero level of skill, the thing we are supposed to know according to most verification systems. We get no brownie points for doing the obvious. But how easy/hard is it "doing" the climate?

Predictability in NWP Models

- We know that smaller the SE, the better the model. However, sufficient hindcasts are usually missing to determine whether the next NWP model has reduced SE relative to previous model(s).

- Even for CFSv2, SE improvement relative to CFSv1 was not a given, because the complete set of hindcasts were made only after a new model configuration was frozen.

- NMME provides another opportunity for comparing SE in different models.

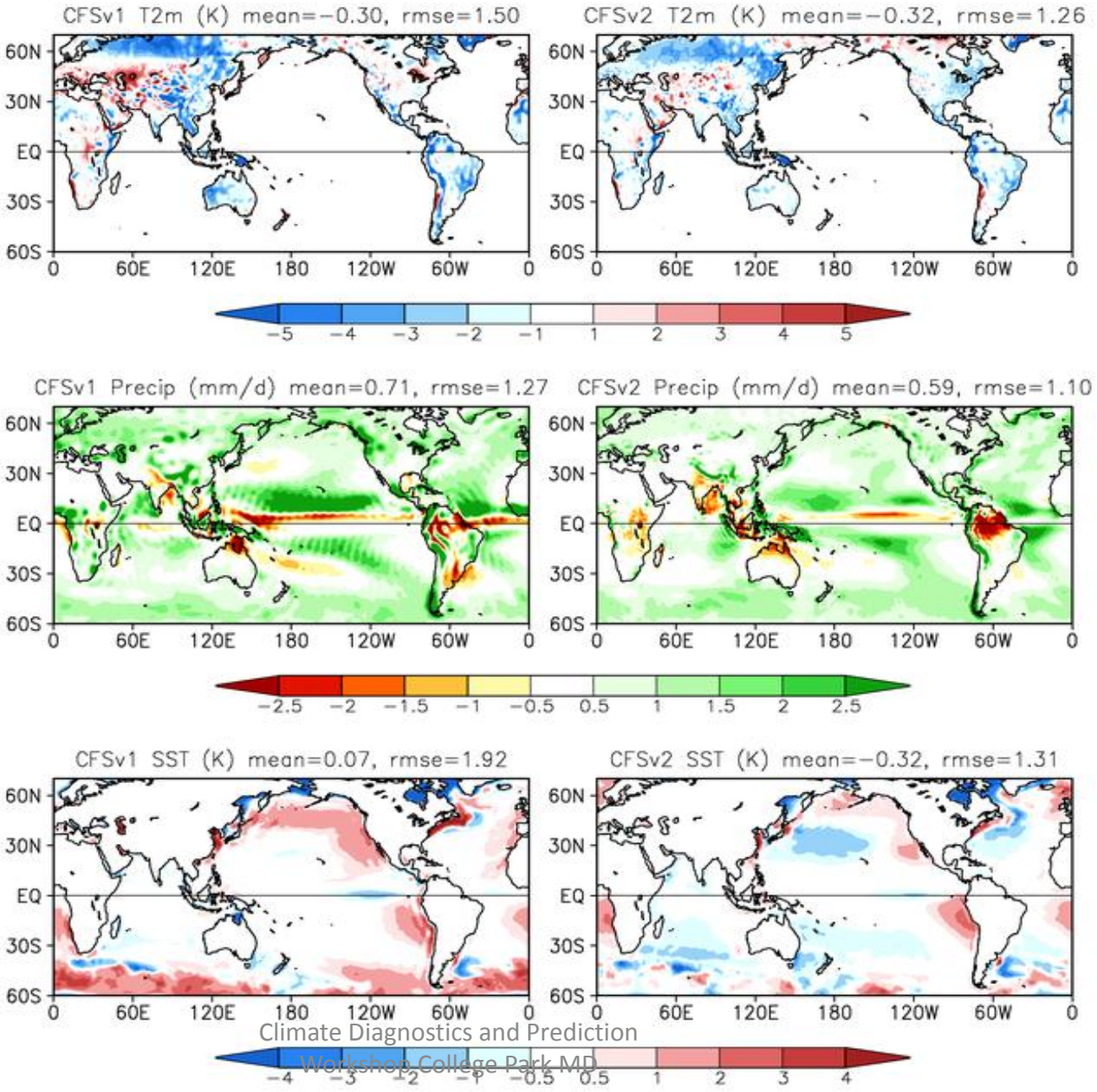
Definitions and Data

- Variables/areas studied: global SST, Precipitation and land only 2-meter temperature.
- GHCN-CAMS (validation for T2m)
- CMAP (validation for Prate)
- QD-OISST (validation for SST)
- 1982-2009 (28 years)
- Common 1.0 degree grid for all models

Hindcast Situation YEAR 2

	Start months available NOW	Period	Members	Arrangement of Members	Lead (months)	Model resident Resolutions: Atmosphere	Ocean	Reference	
NCEP-CFSv1	12	1981-2009	15	1st 0Z +/-2days, 11th 0Z+/-2d, 21st 0Z+/-2d	0-9	T62L64	MOM3L40 0.30 deq Eq	Saha et al 2006	NCEP-CFSv1
NCEP-CFSv2	12	1982-2010	24(28)	4 members (0,6,12,18Z) every 5th day	0-9	T126L64	MOM4 L40 0.25 deg Eq	Saha et al 2010	NCEP-CFSv2
GFDL-CM2.1	12	1982-2010	10	All 1st of the month 0Z	0-11	2x2.5deg L24	MOM4 L50 0.30 deg Eq	Delworth et al 2006	GFDL-CM2.1
CMC1-CanCM3	12	1981-2010	10	All 1st of the month 0Z	0-11	CanAM3 T63L31	CanOM4 L40 0.94 deg Eq	Merryfield et al 2012	CMC1
CMC2-CanCM4	12	1981-2010	10	All 1st of the month 0Z	0-11	CanAM4 T63L35	CanOM4 L40 0.94 deg Eq	Merryfield et al 2012	CMC2
NCAR-CCSM3.0	12	1982-2010	6	All 1st of the month**	0-11	T85L26	POP L40 0.3 deg Eq	Kirtman and Min 2009	NCAR-CCSM3.0
NASA	12	1981-2010	6	1 member every 5th day as CFSv2	0-9	1x1.25deg L72	MOM4 L40 1/4 deg at Eq	Rienecker et al 2008	NASA

Annual Mean Systematic Error (1982–2009) for Lead 3



Saha et al, 2013: The NCEP Climate Forecast System Version 2. Journal of Climate (early online release.) doi: abs/10.1175/JCLI-D-12-00823.1

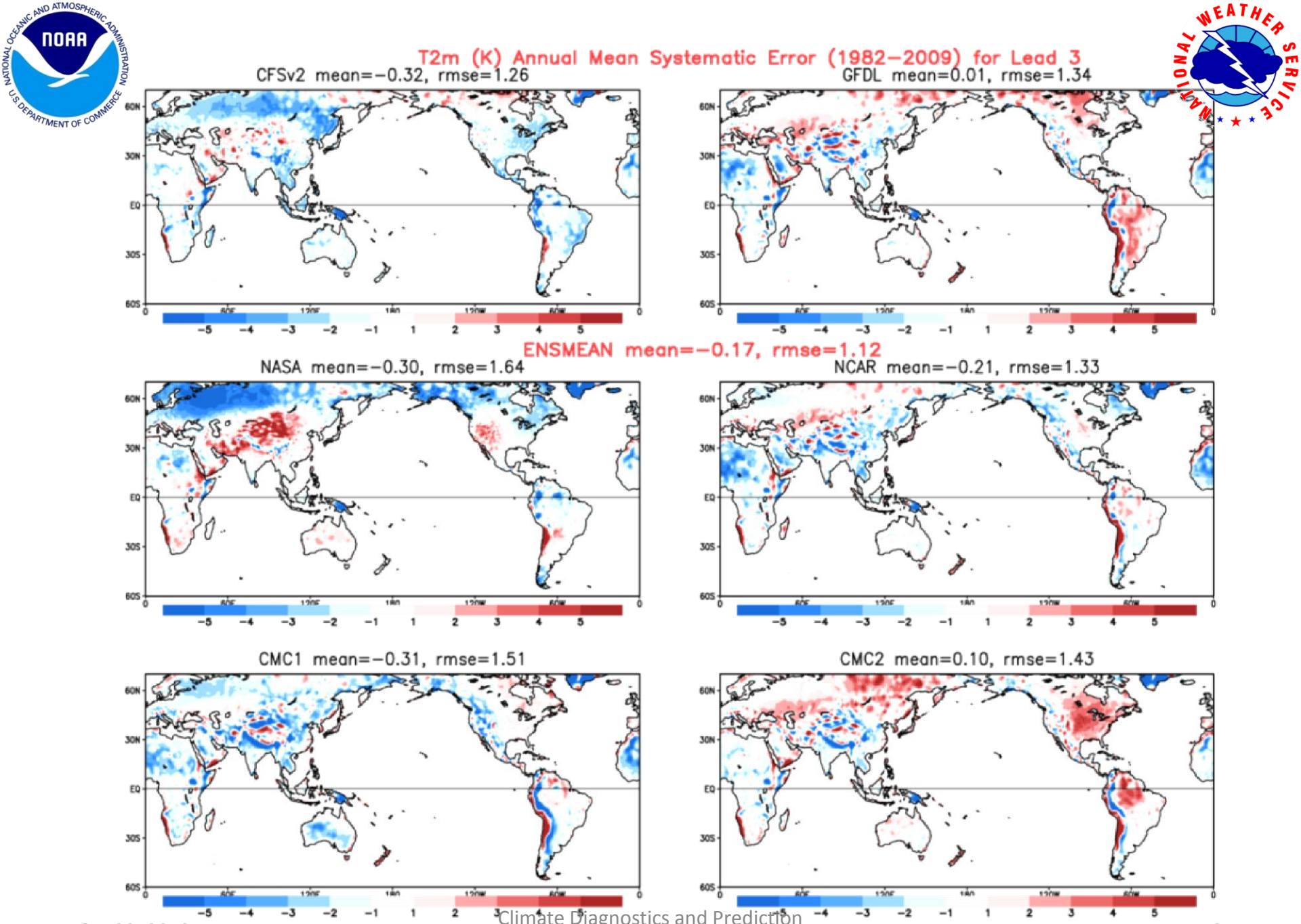
	T2m	T2m	Prate	Prate	SST	SST
Lead	Bias	RMSD	Bias	RMSD	Bias	RMSD
1	-0.26	1.16	0.62	1.10	-0.27	1.26
2	-0.30	1.22	0.60	1.10	-0.30	1.28
3	-0.32	1.26	0.59	1.10	-0.32	1.31
4	-0.34	1.29	0.59	1.10	-0.33	1.34
5	-0.36	1.32	0.58	1.11	-0.36	1.38
6	-0.38	1.35	0.58	1.12	-0.38	1.41
7	-0.41	1.38	0.57	1.12	-0.40	1.45

Systematic Error in CFSv2 , global bias and RMSD for 1982-2009 climo

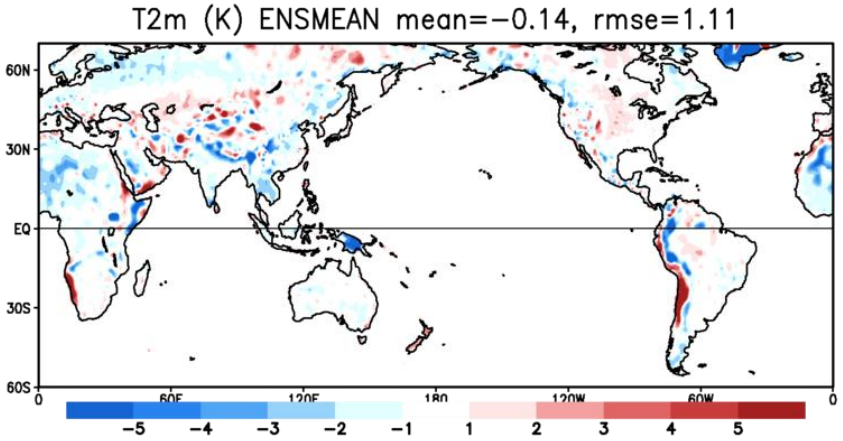
Model is very quickly in a stable SE regime (less than 1 month). This is true for all models.
Lead 3 is representative.
Only small further development of SE later on
No development in prate at all.

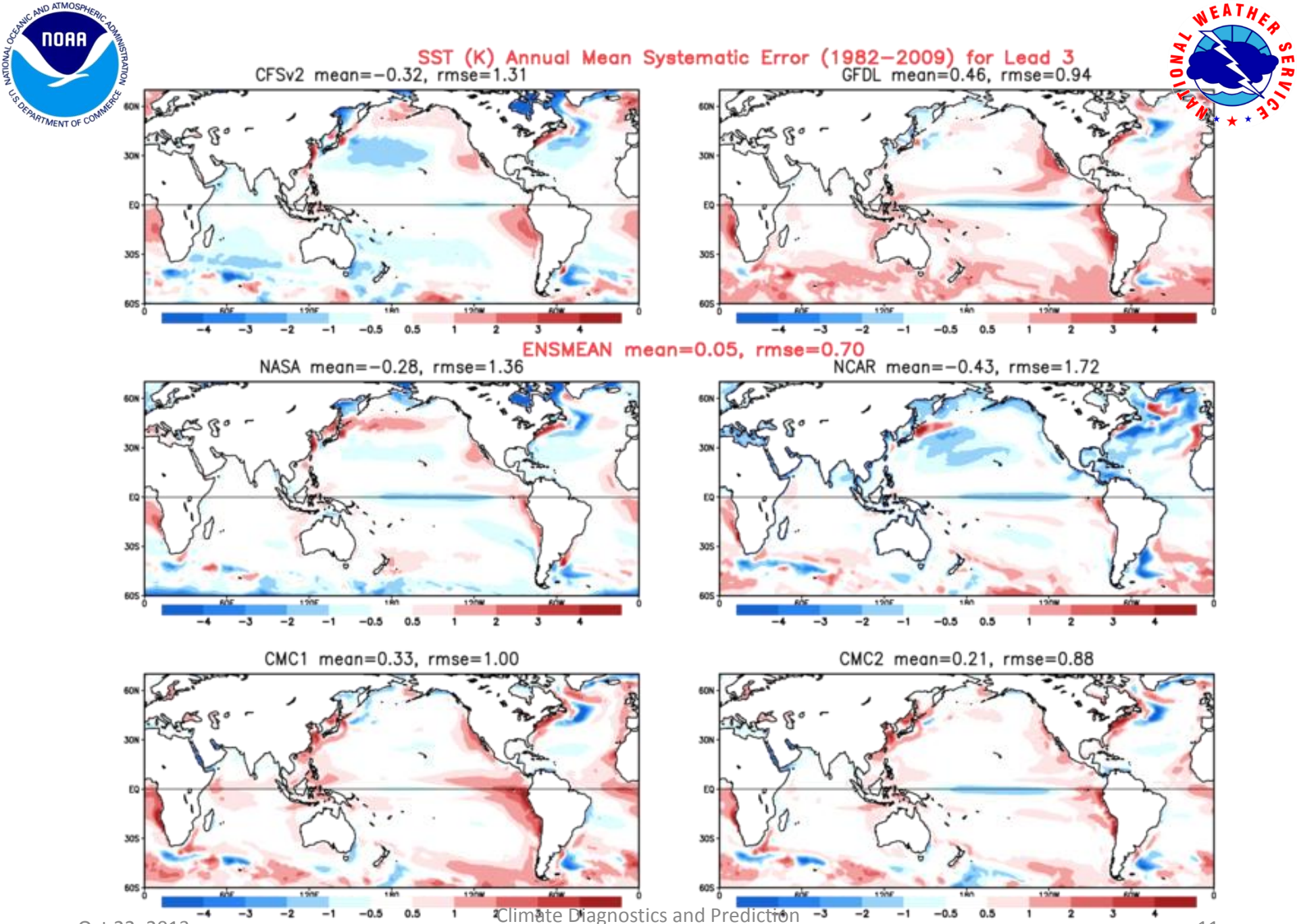
Why lead 3?

- It does not matter much which lead one takes, see next Table first.
- The SE, being relatively independent of lead, does not preclude some very slow development of SE at a much longer time scale.
- For instance, the CFSv2 upper ocean cools very slightly for 15-20 years (in the long CMIP free runs), so increasing CO_2 in the runs does not show warming until after 15 years or so.
- Solution: the difference from a control run with no CO_2 increase will show the correct amount of global warming.

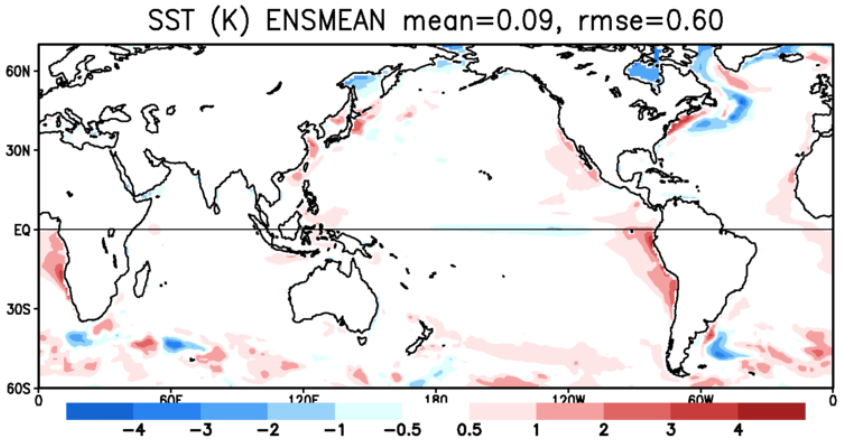

T2m (K) Annual Mean Systematic Error (1982–2009) for Lead 3
CFSv2 mean=−0.32, rmse=1.26
GFDL mean=0.01, rmse=1.34
ENSMEAN mean=−0.17, rmse=1.12
NASA mean=−0.30, rmse=1.64
NCAR mean=−0.21, rmse=1.33
CMC1 mean=−0.31, rmse=1.51
CMC2 mean=0.10, rmse=1.43
60N
30N
EQ
30S
60S
−5 −4 −3 −2 −1 1 2 3 4 5

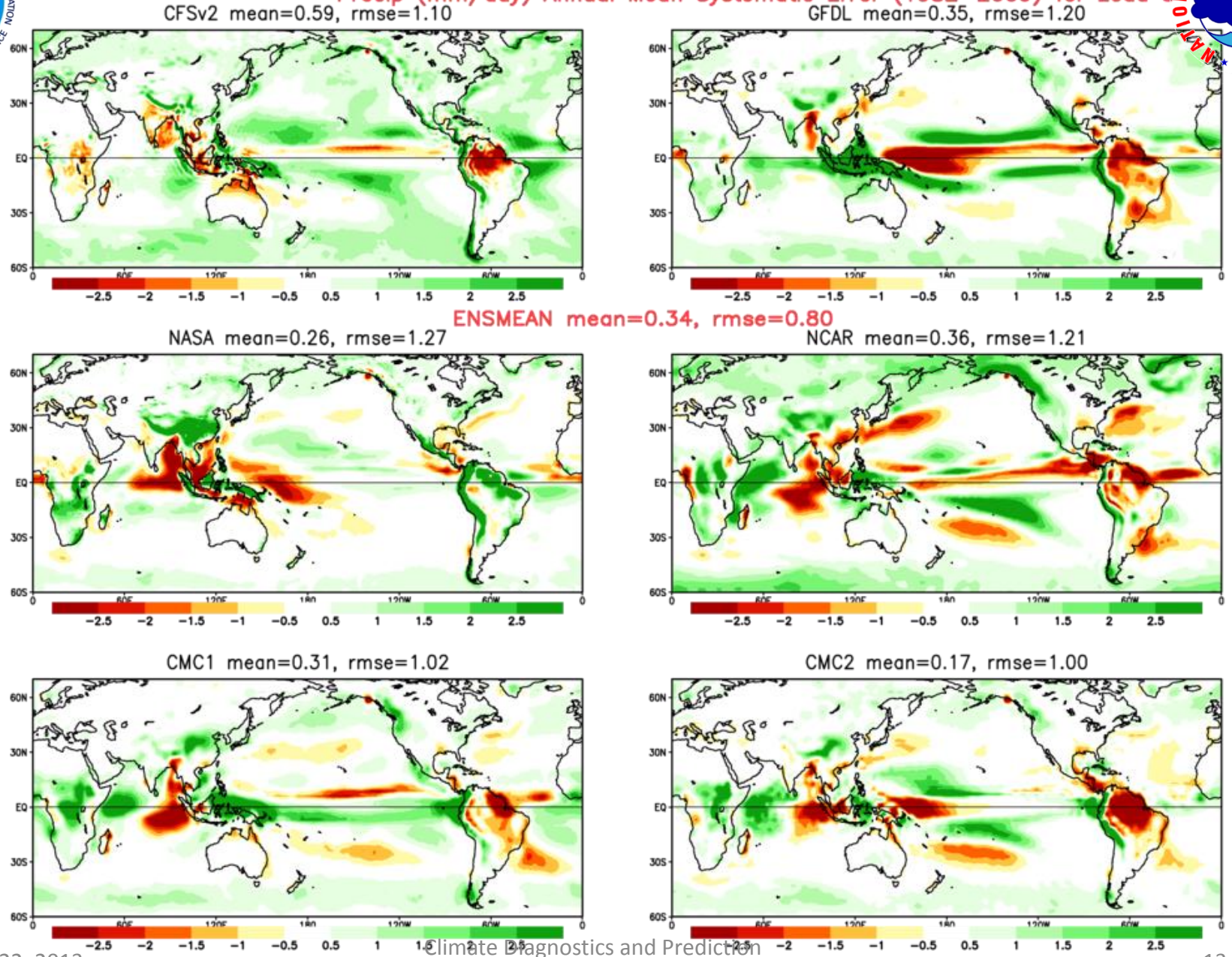
Annual Mean Systematic Error (1982–2009) for Lead 3



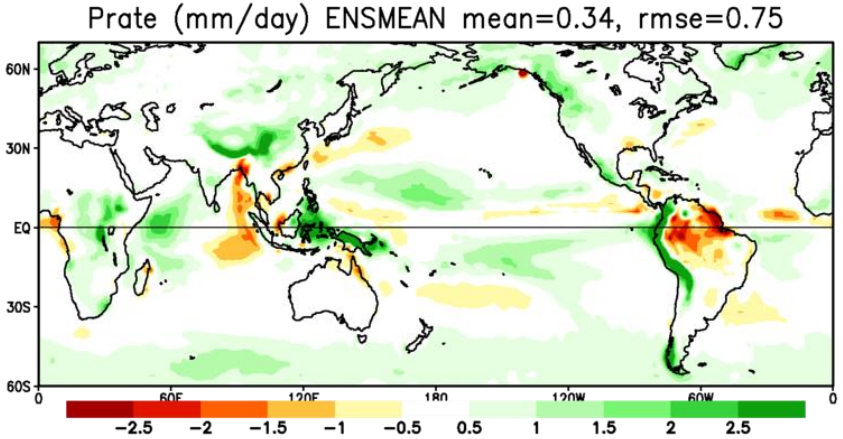

SST (K) Annual Mean Systematic Error (1982–2009) for Lead 3
CFSv2 mean=−0.32, rmse=1.31
GFDL mean=0.46, rmse=0.94
ENSMEAN mean=0.05, rmse=0.70
NASA mean=−0.28, rmse=1.36
NCAR mean=−0.43, rmse=1.72
CMC1 mean=0.33, rmse=1.00
CMC2 mean=0.21, rmse=0.88
60N
30N
EQ
30S
60S
−4 −3 −2 −1 −0.5 0.5 1 2 3 4

Annual Mean Systematic Error (1982–2009) for Lead 3




Precip (mm/day) Annual Mean Systematic Error (1982–2009) for Lead 3
CFSv2 mean=0.59, rmse=1.10
GFDL mean=0.35, rmse=1.20
ENSMEAN mean=0.34, rmse=0.80
NASA mean=0.26, rmse=1.27
NCAR mean=0.36, rmse=1.21
CMC1 mean=0.31, rmse=1.02
CMC2 mean=0.17, rmse=1.00
60N
30N
EQ
30S
60S
0
60E
120E
180
120W
60W
−2.5 −2 −1.5 −1 −0.5 0.5 1 1.5 2 2.5

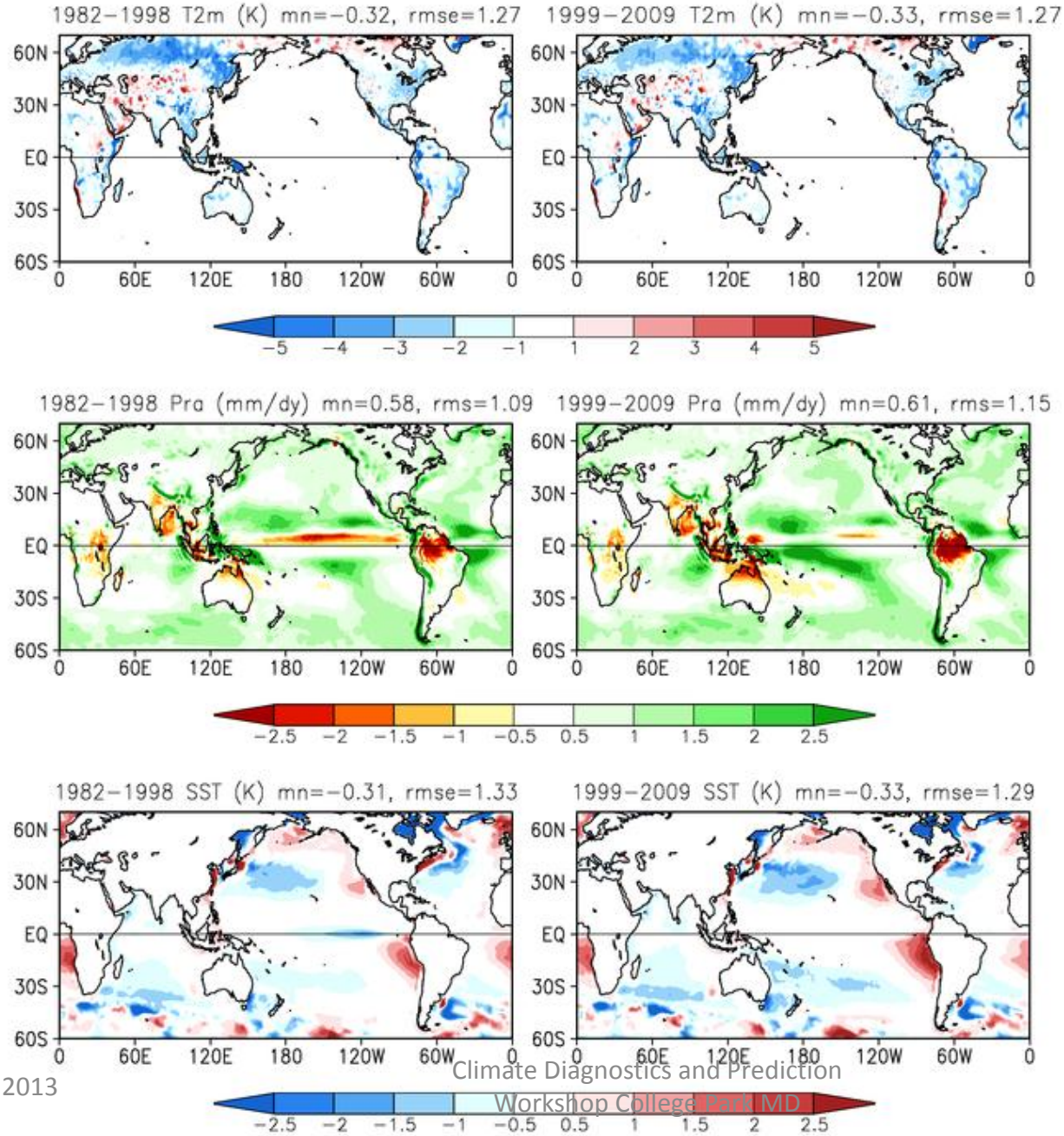
Annual Mean Systematic Error (1982–2009) for Lead 3



Conclusions

- None of the 6 NMME, global coupled ocean atmosphere models, have absurdly large systematic errors (SE) in SST, T2m, and prate.
- SE (annual mean) is reasonably comparable in magnitude among the 6 models.
- One can pick winners and losers, but no model is best for all variables.
- The greater tropical Pacific region is still difficult to model for all three variables. Very different signatures of SE especially in prate.
- SE (annual mean) is smaller in NMME than in most or all of the constituent models. There is a minimization of the error by taking a straight average of all these different models (an old recommendation for NMME).
- SE in July and January (and other individual months) are much larger than the annual mean.
- All maps will be placed on web.

CFSv2 Annual Mean Systematic Error for Lead 3



An entirely different type of evolution